DOCUMENT RESUME

ED 320 415 FL 018 493

AUTHOR Schatz, Henriette F.
TITLE Analyzing English Lexical Elements in the Language of Dutch Immigrants in the United States.
PUB DATE 88
NOTE 11p.; In: York Papers in Linguistics 13. Selected papers from the Sociolinguistics Symposium; see FL 018 472.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Code Switching (Language); *Dutch; *English (Second Language); Foreign Countries; *Immigrants; Language Research; *Linguistic Borrowing; Linguistic Theory; Structural Analysis (Linguistics); Uncommonly Taught Languages
IDENTIFIERS *Language Contact; United States

ABSTRACT

A study investigated the general occurrence of English lexical elements in the recorded speech of 285 Dutch pre-World War I immigrants and their descendants in Massachusetts, Iowa, and Wisconsin. The problems encountered in analyzing a massive data corpus and the methods used to resolve these problems are the focus of the paper. The difficulties of defining code-switching versus borrowing as contrasted with other language contact phenomena are first discussed; next the process of narrowing and operationalizing the characateristics of code-switching is described; and finally it is shown how these characteristics will be applied in computerized form to the data. Results from the "code-switching marker" analysis, used for both single and multi-word items, suggest that the phenomena attending occurrence of English elements can be an important source of information in code-switching analysis, particularly for multi-word elements, and that the categories of analysis must be refined further to determine loan items. Criteria of frequency of occurrence across speakers must be an important additional source of information in code-switching analysis because loans are by definition part of the vocabulary of the speech community as a whole and must occur in the speech of more than a single speaker. Overall, investigation of code-switching and borrowing is seen as complex. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Harlow S.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED320415

ANALYZING ENGLISH LEXICAL ELEMENTS IN THE LANGUAGE OF DUTCH IMMIGRANTS IN THE UNITED STATES

Henriette F. Schatz

(P. J. Meertens Institute for Dialectology
Royal Netherlands Academy of Arts and Sciences)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

1.0 Introduction

In this paper I will be discussing an investigation of the speech of Dutch immigrants in the United States. I will do so on the basis of some data from a very large corpus of tape-recorded language material, collected in the United States in 1966 (cf. Daan, 1987). This data collection covers approximately 75 hours of recordings of 285 different speakers from nine, mostly nineteenth-Century, settlements in Massachussetts, Iowa, and Wisconsin. All these speakers either have immigrated before World War I or are the children, and -- in some cases the grandchildren -- of immigrants from that period.

In choosing the speakers, the point of departure was that the speakers themselves, their parents, or their grandparents must have immigrated before the major social changes after World War I had taken place in the Netherlands, because these changes had a profound effect on the use and the nature of dialects in the Netherlands. The recordings contained free conversations, mostly on topics of interest to the speakers themselves, but sometimes directed by the interviewers' questions. The setting was relatively formal for a free conversation: the interviewers were two perfect strangers who elicited Dutch or dialect from speakers who often hardly used these language-varieties in their everyday life.

The original purpose of collecting the American Corpus was "to document the history of the Dutch language, its dialects, and the Frisian language in the United States" (KNAW, 1966:623). Expectations were that by using this method of data collection it would be possible to elicit and record data from older stages of the Dutch dialects, supposedly preserved in the American immigrant communities. These expectations were based on the notion that the immigrants would not have been subject to the same influences that conditioned dialect change in the Netherlands and would, therefore, have preserved older forms of the dialects. Having such older stages of the language on tape, as opposed to having them available only in written form, could provide a "dialect museum", as it were. Legitimate as this notion might have been at the time, it disregards completely that there could have been a host of other factors conditioning changes in the immigrants' Dutch dialects. As a result, the language corpus that was collected turned out not to be particularly useful for the description of older dialect forms. On the other hand, it does provide an extremely valuable source of recorded material for the investigation of a young language contact situation resulting from immigration. For this reason, a large-scale research project is currently underway to study various linguistic aspects of this material (cf. Schatz, 1987; Smits, 1988).

2.0 Language contact phenomena

The original purpose of my research was to study code-switching in the language of Dutch immigrants, on the basis of natural language data. In the course of my investigation of the available material I have gradually changed my focus to the general occurrence of English lexical elements in the immigrants' speech. In this paper I will concentrate on some of the problems I have encountered in analyzing a massive data corpus for phenomena of language contact and my ways of coping with these problems. I will first discuss the problems of defining code-switching and borrowing, as opposed to other language contact phenomena. Then I will show how I have narrowed down the characteristics of code-

York Papers in Linguistics 13 (1989) 307-316
© the author

BEST COPY AVAILABLE

FL018493

Paper presented at the Sociolinguistics Symposium (7th, York, England, April 14-16, 1988).

switching and how I have operationalized them for use on my own data corpus. Finally, I will apply them -- in the form of codes in a computerized data entry form -- to a 2317 item test corpus of English lexical elements, in order to ascertain the feasibility of this method of analysis.

3.0 The available data

The more practical reason for a change of focus from code-switches, in a strict sense, to English lexical elements in general, was the available data. A direct result of expecting to record "earlier stages of Dutch" was that the tape-recorder was switched off at any time a speaker used more than a few English words. Audio-tape was still an expensive commodity at the time and, in addition, the interviewers regularly admonished the informants to speak only Dutch or dialect, an elicitation technique which is not exactly conducive to recording a large corpus of natural code-switches.

The interviewers' intrusions in the otherwise free conversations obviously influenced where and when informants switched from English to Dutch, as well as the number of English elements in the total language corpus. It is all the more surprising, then, that the corpus yielded some 12000 instances of English elements in the immigrants' spoken Dutch, counting only the data for the almost one hundred first generation immigrant speakers. The sheer numbers show that the use of English elements must depend on other factors than conscious control, and my eventual purpose is to determine what controlling factors there are -- linguistic and otherwise -- in the use of English by these immigrants.

4.0 Definitions of code-switching

A more linguistically oriented reason for investigating English elements in general is the fact that the more I attempted to focus on code-switching, the more elusive the concept became. It has proven quite difficult to disentangle code-switching, code-mixing, code-changing, borrowing, nonce-borrowing, lexical borrowing, grammatical interference, and transference. Often these terms are ill-defined, interchangeable, or even in conflict with one another. The concepts covered by them are all used in the theoretical and terminological discussions in literature about language contact phenomena, and the resulting confusion may very well influence the results of research (cf. Clyne, 1987:741).

My problem in formulating a tight definition for code-switching, borrowing, and a number of related aspects of language contact for the purpose of data analysis is surely familiar to many linguists. On the basis of some recent literature, I attempted to narrow down the factors that distinguish code-switching from other language contact phenomena. In the course of this literature review, I have become convinced that the characteristics that are generally employed to define code-switching and borrowing more or less cover all the distinctions between language contact phenomena as they may be found in the literature under a variety of names.

5.0 Defining borrowing versus code-switching

Reviewing a fairly arbitrary choice of some recent articles on code-switching and related phenomena (Aguirre, 1985; Appel en Muysken, 1987; Bentahila and Davis, 1983; Berk-Seligson, 1986; Crama and van Gelderen, 1984; Clyne, 1987; Gingrás,1974; Gumperz,1976; Kumar, 1986; Muysken, 1987; Moravcsik, 1978; Pfaff, 1979; Poplack, 1980 and 1987; Sankoff and Poplack, 1985; Sankoff, Poplack and Miller, 1988; Sankoff, Poplack and Vanniarajan, 1985; Shaffer, 1975; Sridhar and Sridhar, 1980), as well as some earlier literature on the subject (Bloomfield, 1933; Haugen, 1950 and 1972; Weinreich, 1953), provided me with a choice of authors' definitions for code-switching.

Characteristic	Borrowing	Code-Switching	Data US
a. single word	+		+
b. both single and multi-word	+	+	+
c. multi-word element		+	+
phonological integration in L1	+		+
no phonological integration in L1		+	+
morphological integration	+		+
no morphological integration		+	+
grammatical integration	+		-
no grammatical integration		+	-
no simultaneous integration on all levels		+	+
mono- and bilingual speakers	+		-
bilingual speakers only		+	+
lexical need	+		-
treated as part of L1 lexicon	+	+	-
relatively frequent word	+		+
widely used in speech community	+		-
recognition, acceptance, approval in speech community	+		-
only from L2 to L1	+		-
only lex. items from L1 in L1 element and from L2 in L2 element		+	-
L1 equivalent exists		+	+
L1 equiv. always used in speech in speech community		+	-
L1 equiv. known in sp. comm.		+	-
Tags, interjections, idiomatic expressions, function words, yes, no, hesitations		+	+

Table 1. Defining characteristics of Borrowing and Code-Switching in the literature on code-switching and borrowing.

In addition, I found definitions for other terms, such as code-mixing, lexical borrowing, nonce-borrowing, transference, and grammatical interference. In effect, they often seemed to cover the same ground as code-switching or borrowing under a different name. I decided to focus on the distinction between code-switching and borrowing and listed all phenomena which, according to their defining characteristics in the literature, could be termed code-switching or borrowing, even if they were discussed under some other name. The definitions are given in Table 1, which also lists whether or not they can be tested on my data from the United States.

6.0 Analysis of the test corpus

From the12000 item data collection I gathered a test corpus comprising 2317 English elements. First I checked against two other data collections of spoken language whether this test corpus could be considered representative for the whole data collection (cf. section 7.0). Then I tried out the list of characteristics of code-switching and borrowing on the test corpus to find out whether such characteristics could profitably be used in the analysis of real language data. In order to do so, I operationalized them in a coding system and a data entry form. All of the items from the test corpus were then coded in a number of categories.

Some of the characteristics could not be tested on the Dutch data, simply because the information was not available. For example, since the data are all from bilinguals, differences between bilingual and monolingual speakers cannot be investigated. Information about the speech community in general, or information on speaker opinions, speaker awareness, or speaker knowledge could not be investigated, simply because it was not elicited by the interviewers. In Table 1, such characteristics are marked minus in the column 'Data US'.

The distinction between single and multi-word items is considered important in much of the literature on code-switching, so I distinguished between single and multi-word items.

Because many cross-sentential switches had been controlled or cut short by the elicitation method, I concentrated on intrasentential items and a distinction was made between inter- and intrasentential items.

The question of grammatical integration is also often discussed in the literature, either in general, or in particular reference to phonological, morphological or syntactic phenomena. For this reason a code was given to each data entry for which information was available about the degree of integration of the English element.

Whether or not another English element occurs els e where in a speaker's sentence can also yield information about code-switching behavior, so that information was also recorded.

I am convinced that the context of an element is very important. It is needed, for example, to obtain information about phenomena that often accompany code-switches, such as repetitions, hesitations, translations, reported speech, false starts, or meta-linguistic commentaries. For this reason the data themselves were entered in two ways: the string or single English element alone and the same element in its context.

Special remarks about a particular English element were also noted. Each data item in the corpus remains uniquely identifiable by tape number, informant number, and page number of the transcript, so if a larger context is needed for any reason, it is possible to go back to the transcripts, or even to the recordings themselves.

As far as grammatical categories are concerned, single items were coded for part of speech, while multi-word items were coded for syntactic category. In additon, I drew up a fairly exhaustive list of various phenomena occurring in the speech of bilinguals when code-switching is taking place, based on the literature about the subject (particularly Poplack 1980 and1987 and Berg Seligson 1986). These phenomena include repetitions,

hesitations, false starts, translations, metalinguistic comments, etc., and I consider them characteristic for code-switching. All of these attendant phenomena were coded in a separate category which I termed 'code-switch markers'. Other information, for example on morphological or phonological integration of the English element or the occurrence of Dutch/English compounds, was coded under the category 'other information'.

A special category is formed by the list of what I have termed 'fillers'. These words or phrases are often used as a matter of habit by many of the speakers. I assigned them a grammatical code, like all other items, but, in addition, I coded them in a separate category because they differ so much from other English lexical elements, both in their frequency of occurrence and in their use in otherwise completely Dutch discourse. In an analysis of English elements, these items might be viewed as belonging to a separate category and occurring under special conditions that set them apart from regular English items. Since most of the fillers are to be found in the category of conjunctions, the number of occurrences in this word-class is relatively large.

7.0 Some quantitative results

In analyzing English elements from the point of view of code-switching, I focused on single-word items first, because they form the majority of my data, due to the limitations of the data collection itself.

Out of the test corpus of 2317 data items, 1471 (63.5%) were single English elements. Broken down by parts of speech (figure 1) the distribution shows that nouns are by far the largest category (24.6%). The next important category is that of conjunctions (20.3%), followed by adverbs (18.4%), verbs (including auxiliaries) (13. 6%), pronouns (5.3%), adjectives (2%), and prepositions (0.9%), while negations, affirmatives, determinters, and numerals form the remainder (14.9%). Excluding fillers, which make up a large proportion of the conjunctions, this category would only amount to 1.8% of the single elements.

In order to see whether or not this test corpus was -- at least to some extent -- a representative collection of English elements, I checked my results against the ones reported by Poplack (1980) for an English/Spanish contact situation and against the normal distribution of word classes for Standard Dutch (De Jong, 1979) in the expectation that I would find greater similarities between the two sets of immigrant data than between the word class distribution in Standard Dutch and the distribution of English elements in Dutch immigrant speech.

Comparing the distribution for single-word elements across word classes, the order of frequency is strikingly similar, as can be seen in Table 2. In the results reported by Poplack, single-word and multi-word elements were not considered as separate categories, so for the sake of comparison the percentage of the total number of English elements (including multi-word items) is also given for the Dutch immigrant data.

The 846 (36.5%) multi-word items in the test corpus were assigned to different syntactic categories. Their distribution in descending order is as follows: Sentence (21.7%), Indep. Clause (20%), Phrase (13.6%), Subject NP (12.6%), Object NP (12.6%), Interjection (6.9%), VP (5.6%), Comp (2.6%), Coord. Clause (3%), Subord. Clause (0.8%), Tag (0.6%). A comparison with Poplack's results for the multi-word elements yields fewer similarities for the two corpora, as can be seen from Table 3. Nevertheless, the general tendencies seemed similar enough to consider my test corpus representative for a large body of code-switching data.

	N	Conj.	Adv.	V	Pro.	Adj.	Prep.	
Schatz 1988	24.6	20.3*	18.4	13.6	5.3	2	0.9	% of sing. el's
	15.6	12.9*	11.7	8.6	3.4	1.3	0.6	% of total el's
Poplack 1980	9.5	2.7**	2.6	1	***	0.8	0.1	% of total el's

* Conj. include 18.2% fillers
** Conj. include rel. pronoun
*** No data available

Table 2. Comparison across word classes of non-native elements in the speech of Dutch and Spanish immigrants in contact with English.

	Schatz 1988		Poplack 1980
	% of sing. el's	% of total el's	% of total el's
Sentence	21.7	7.9	20.3
Indep. Clause	20	7.3	4.3
Phrase	13.6	5	5.1
Subject NP	12.6	4.6	3.8
Object NP	12.6	4.6	7.6
Interj./ Tag	7.5	2.7	26.8
VP	5.6	2	2.2
Coord. Clause	3	1.1	*
Comp.	2.6	0.9	*
Subord. Clause	0.8	0	4.1

* no data available

Table 3. Comparison across syntactic categories of non-native elements in the speech of Dutch and Spanish immigrants in contact with English.

The following chart (figure 1.) shows the distribution across word classes of single English elements in the speech of the Dutch immigrants as it compares with the distribution of word classes in spoken Standard Dutch (De Jong, 1979). It is clear that the distribution of single English elements in the speech of the Dutch immigrants shows a greater similarity with that for Spanish immigrant speakers than for Standard Dutch speakers. This would

suggest that -- at least for language contact with English -- the structure of English dictates to some extent which word classes are more likely to be the target for code-switching or borrowing.

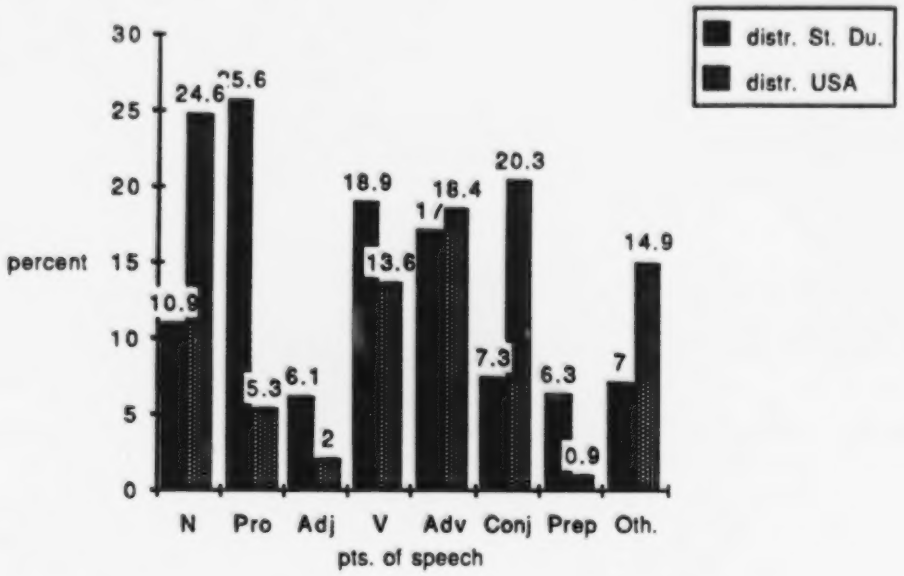

Figure1. Distribution across word classes of spoken Standard Dutch and English elements in the speech of Dutch immigrants in the United States.

As for the question of code-switching, I carried out a brief analysis of the distribution of 'code-switch markers' across single and multi-word items, in order to see whether these phenomena occuring 'around' English elements would help to identify code-switches. Out of the 1417 single items, 532 (36.2%) were fillers, which I did not take into consideration, so I was left with 885 single items. Only 116 (7.9%) of these were marked as code-switches: 33 hesitations, 18 false starts, 17 metalinguistic comments, 16 translations, 15 repetitions, 7 explanations, 5 proper names, and 3 other flags. For the time being, I also counted the 31 homophonous diamorphs as code-switches, because in effect they may belong to either code and it is difficult, if not impossible, to decide to which language these utterances must be assigned.

In all, 147 (16.6%) of the 885 remaining single elements could be assigned a code-switch marker, which implies that single items cannot clearly be defined on the basis of these markers. In judging these exploratory results it must be kept in mind, though, that not all single English elements will be code-switches: a substantial percentage of them should be English loans. They would not be marked for code-switching in any event, but the problem is that it is not possible to determine what percentage of single elements must be considered loans.

For the multi-word items the results were more informative. Out of the 846 multi-word items, 52 were interjections (6.1%) and 101 (11.9%) were fillers. These were excluded from the analysis. Of the remaining 693 multi-word items 275 (39.6%) were marked as code-switches in one way or another (71 hesitations, 35 false starts, 22 metalinguistic comments, 16 translations, 33 repetitions, 8 explanations, 27 proper names, 45 other flags, 5 word-search problem, 4 changed interlocutor). The 36 homophonous diamorphs, were also counted as code-switches, bringing the total of 'marked' switches upt to 311 (44.9%), which is a substantial part of the multi-word items.

8. Conclusion

From the 'code-switch marker' analysis for single as well as multi-word items it is clear that the phenomena attending the occurrence of English elements can be an important source of information in the analysis of code-switching, be it more so for multi-word elements than for single words. It is also clear that the categories of analysis must be refined further, in order to determine which items must be considered loans.

Criteria of frequency of occurrence across speakers must be an important additional factor in this analysis, since loans are by definition part of the vocabulary of the speech community as a whole and must therefore occur in the speech of more than a single speaker. If a single English element occurs across the speech community with a fair degree of frequency, it can safely be considered a loan. This also applies to elements occuring frequently in the speech of one particular speaker. However, in the case of infrequent items such a decision is not so easily made, since it is in fact a matter of chance whether or not an item will occur. In such cases other linguistic information as well as sociolinguistic data may provide clues for analysis.

Whatever the outcome of further work on this data corpus will be, this brief analysis has shown that the investigation of code-switching and borrowing is by no means a simple and straightforward matter. Since the large data corpus under investigation here is more or less representative for the speech community and accessible by computer it will be possible to carry out a combined descriptive and quantitative analysis of switched language elements on the basis of these data. In my opinion, this type of careful analysis should precede any claims about factors constraining code-switching, both on a formal linguistic and on a functional level.

REFERENCES

Aguirre, A., 1985. An Experimental Study of Code-alternation. *International Journal of the Sociology of Language* , 53: 59-82.

Appel, R. en P. Muysken, 1987. *Language Contact and Bilingualism*. London.

Bentahila, A. en E. Davies, 1983. The Syntax of Arabic-French Code-Switching. *Lingua*, 59: 301-330.

Berk-Seligson, S., 1986. Linguistic Constraints on Intrasentential Code-Switching: A study of Spanish/Hebrew Bilingualism. *Language in Society*, 15, 3: 313-348.

Bloomfield, L., 1933. *Language*. New York.

Clyne, M., 1987. Constraints on Code Switching: How Universal Are They?. *Linguistics*, 25:739-764.

Crama, R. and H. van Gelderen, 1984. *Structural Constraints on Code-Mixing*. Amsterdam: Instituut voor Algemene Taalwetenschap.

Daan, J., 1987. *Ik Was te Bissie... Nederlanders en hun Taal in de Verenigde Staten*. Zutphen.

De Jong, E., 1979. *Spreektaal: Woordfrequenties in Gesproken Nederlands*. Utrecht.

Gingrás, R., 1974. Problems in the Description of Spanish/English Intra-Sentential Code-Switching. In: G. Bills (Ed.). *Southwest Areal Linguistics*. San Diego, Ca.

Gumperz, J.J., 1976. The Sociolinguistic Significance of Conversational Code-Switching. In: J. Cook-Gumperz, (Ed.). *Papers on Language and Context*. Berkely, Ca.

Haugen, E., 1950. The Analysis of Linguistic Borrowing. *Language*, 26:210-232.

Haugen, E., 1972. *The Ecology of Language*. Stanford, Ca.

KNAW, 1966. *Jaarboek der Koninklijke Nederlandse Akademie van Wetenschappen 1966/1967*. Amsterdam: Noord Hollandsche Uitgevers Maatschappij.

Kumar, A., 1986. Certain Aspects of the Form and functions of Hindi-English Code-Switching. *Anthropological Linguistics*, 28:195-205.

Moravcsik, E., 1978. Universals of Language Contact. In: J. Greenberg. *Universals of Human Language*. Stanford, Ca., 95-122.

Muysken, P., 1987. Neutrality in Code-Switching. In: *Eigen en Vreemd: Handelingen van het 39e Filologencongres*. Amsterdam, Vrije Universiteit, 359-373.

Nortier, J.M. and H.F. Schatz, 1988. Van Eenwoordwisseling naar Ontlening: Een vergelijkend onderzoek. In: *Forum der Letteren*, 29,3.

Pfaff, C., 1979. Constraints on Language Mixing: Intra-Sentential Code-Switching and Borrowing in Spanish-English. *Language*, 55, 1979: 291-318.

Poplack, S., 1980. Sometimes I'll Start a Sentence in Spanish Y TERMINO EN ESPANOL: Toward a Typology of Code-Switching. *Linguistics*, 18:581-618.

Poplack, S., 1987. Contrasting Patterns of Code-Switching in Two Communities. In: Erlang Wande e. a.. *Aspects of Multilingualism*. Uppsala.

Poplack, S., D. Sankoff, en C. Miller, 1988. 'The Social Correlates and Linguistic Processes of Lexical Borrowing and Assimilation'. *Linguistics*, 26:47-104.

Sankoff, D. and S. Poplack, 1985. Adposition and Article in Finnish-English Code-Switching. Unpublished Ms., Université de Montréal.

Sankoff, D., S. Poplack, and S. Vanniarajan, 1985. The Case of Nonce-Loan in Tamil. Unpublished Ms., Université de Montréal.

Schatz, H.F, 1987. De Taal van Emigranten in de Verenigde Staten. *Taal en Tongval*, 39: 97-108.

Shaffer, D., 1975. The Place of Code-Switching in Linguistic Contacts. In: *Linguistic Association of Canada, Second Annual Forum*, 487-496.

Smits, C., 1988. Taalverlies in de U.S.A.: Generalisaties in Werkwoordsvervoegingen bij Nederlandse Immigranten in Amerika. Unpublished Ms.

Sridhar, S. and K. Sridhar, 1980. The Syntax and Psycholinguistics of Bilingual Code-Mixing. In: *Studies in the Linguistic Sciences*, 10,1.

Weinreich, U., 1953. *Languages in Contact*.. New York.